13 Rec'd PCT/PTO 02 DEC 2001
PCT/CA00/00716

09/980706

**Title:** A marker specific for *Escherichia coli* serotypes O157:H7; O157:NM and O55:H7

## FIELD OF THE INVENTION

This invention relates to a novel DNA marker specific for *E. coli* serotypes O157:H7, O157:NM and O55:H7 and the use of the marker in developing assays to detect these serotypes of E. coli in a sample.

## BACKGROUND OF THE INVENTION

*E. coli* O157:H7 is a food-borne human pathogen causing a spectrum of diseases including diarrhea, hemorrhagic colitis (HC) and hemolytic uremic syndrome (HUS). Development of nucleic-based assays for rapid detection of *E. coli* O157:H7 has been challenging due to a lack of specific DNA markers for this organism. The reported DNA markers for *E. coli* O157:H7 have one or more of the following limitations or drawbacks: being non-specific, so that multiplex PCR assays need to be used; and containing only one base pair mutation, limiting their use in assay development. Consequently, there is a need in the art to provide novel markers for E. coli O157:H7 and related serotypes.

## SUMMARY OF THE INVENTION

The present inventors have prepared a novel marker specific for E. coli serotypes O157:H7, O157:NM and O55:H7. The DNA sequence of the marker has a total of 1583 nucleotides and has no significant homology to any known DNA sequences. Accordingly, in one aspect, the present invention provides an isolated nucleic acid sequence comprising the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof.

The marker of the invention can be used to develop probes or primers that can be used to detect E. coli serotypes O157:H7; O157:NM and O55:H7 in a sample. In particular, the marker has been demonstrated to be useful in designing primers for PCR assays for specific detection of the *E. coli* serotypes. As the specific DNA sequences of this invention are not homologous to previously known sequences, various specific PCR assays

can be developed with only one primer pair. In addition, other nucleic acid-based assays such as DNA chip or biosensor assays can also be developed without the restriction in using a very limited region of a marker. Therefore, the invention can be used to develop nucleic acid based assays to detect the E. coli serotypes O157:H7, O157:NM and O55:H7.

Other features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

## BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in relation to the drawings in which:

Figure 1 and SEQ.ID.NO.:1 shows the nucleotide sequence of the 1583 bp marker of the invention.

Figure 2 and SEQ.ID.NO.:2 shows the nucleotide sequence of a 360 bp fragment of the marker.

Figure 3 is a gel electrophoresis of PCR products amplified using the PCR assay from DNA preparations of representative *E. coli* serotypes: lanes 1, 2, O157:H7; lanes 3, 4, O157:NM; lanes 5 - 9, O145:NM; lanes 10, 11, O55:H7; lane 12, O26:H11; lane M, 100 bp DNA ladder.

## DETAILED DESCRIPTION OF THE INVENTION

### I. MARKER

As hereinbefore mentioned, the present inventors have prepared a novel DNA marker specific for E. coli serotypes O157:H7, O157:NM and O55:H7.

The marker was identified and isolated using a relatively new marker technology, fluorescent amplified fragment length polymorphism (FAFLP) and DNA sequencing. The DNA sequence of the marker has a

total of 1583 nucleotides and has no significant homology to any known DNA sequences.

Accordingly, in one aspect, the present invention provides an isolated nucleic acid molecule comprising the sequence shown in Figure 1 (SEQ.ID.NO.:1), wherein T can also be U, or a diagnostic fragment thereof.

The term "isolated" refers to a nucleic acid molecule substantially free of cellular material or culture medium when produced by recombinant DNA techniques or chemical precursors when chemically synthesized. The term "nucleic acid" includes deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) and can either be double stranded or single stranded.

The term "diagnostic fragment" means any fragment of the marker shown in Figure 1 (SEQ.ID.NO.:1) that is useful in a diagnostic assay to detect E. coli serotypes O157:H7; O157:NM and O55:H7. The diagnostic fragment includes fragments that can be used as primers in PCR assays and fragments that can be used as probes in detection assays. The diagnostic fragments of the invention will not cross react with bacteria other than E. coli serotypes O157:H7; O157:NM and O55:H7.

The invention also includes (1) a nucleic sequence that is complimentary to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (2) a nucleic sequence that can hybridize to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (3) a nucleic acid sequence which has substantial sequence homology to (1) or (2); and (4) a nucleic acid sequence which is an analog of the nucleic acid sequences of (1) to (3). Such sequences are also useful in developing diagnostic assays to detect E. coli serotypes O157:H7; O157:NM and O55:H7.

The term "a nucleic acid sequence which has substantial sequence homology" means a nucleic acid sequence which has slight or inconsequential sequence variations from the sequence shown in Figure 1 (SEQ.ID.NO.:1). For example, one skilled in the art can appreciate that variations can be made to the sequence of Figure 1 (SEQ.ID.NO.:1) and still

permit the use of the sequence in detecting the E. coli serotypes of the invention. In addition, natural variations may exist in the sequence of certain isolates which may be attributable to local mutations or structural modifications. Nucleic acid sequences having substantial homology include nucleic acid sequences having at least 85%, preferably 90-95% identity with the nucleic acid sequence as shown in Figure 1 (SEQ.ID.NO.:1).

The term "a nucleic acid sequence that can hybridize" means a nucleic acid sequence that can hybridize to a nucleic acid sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragments thereof under hybridization conditions, preferably stringent hybridization conditions. Appropriate stringency conditions which promote DNA hybridization are known to those skilled in the art, or may be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the following may be employed: 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C. The stringency may be selected based on the conditions used in the wash step. For example, the salt concentration in the wash step can be selected from a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be at high stringency conditions, at about 65°C.

The term "a nucleic acid sequence which is an analog" means a nucleic acid sequence which has been modified as compared to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof wherein the modification does not alter the utility of the sequence (i.e. to detect E. Coli serotypes O157:H7; O157:NM and O55:H7 or to develop probes, primers or microchips to detect the serotypes) as described herein. The modified sequence or analog may have improved properties over the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof. One example of a modification to prepare an analog is to replace one of the naturally occurring basis (i.e. adenine, guanine, cytosine or thymidine) of the sequence shown in Figure 1 (SEQ.ID.NO.:1) with a modified base such as such as xanthine, hypoxanthine, 2-aminoadenine, 6-

methyl, 2-propyl and other alkyl adenines, 5-halo uracil, 5-halo cytosine, 6-aza uracil, 6-aza cytosine and 6-aza thymine, pseudo uracil, 4-thiouracil, 8-halo adenine, 8-aminoadenine, 8-thiol adenine, 8-thiolalkyl adenines, 8-hydroxyl adenine and other 8-substituted adenines, 8-halo guanines, 8-amino guanine, 8-thiol guanine, 8-thiolalkyl guanines, 8-hydroxyl guanine and other 8-substituted guanines, other aza and deaza uracils, thymidines, cytosines, adenines, or guanines, 5-trifluoromethyl uracil and 5-trifluoro cytosine.

Another example of a modification is to include modified phosphorous or oxygen heteroatoms in the phosphate backbone, short chain alkyl or cycloalkyl intersugar linkages or short chain heteroatomic or heterocyclic intersugar linkages in the nucleic acid molecule shown in Figure 1 (SEQ.ID.NO.:1). For example, the nucleic acid sequences may contain phosphorothioates, phosphotriesters, methyl phosphonates, and phosphorodithioates.

A further example of an analog of a nucleic acid molecule of the invention is a peptide nucleic acid (PNA) wherein the deoxyribose (or ribose) phosphate backbone in the DNA (or RNA), is replaced with a polyamide backbone which is similar to that found in peptides (P.E. Nielsen, et al Science 1991, 254, 1497). PNA analogs have been shown to be resistant to degradation by enzymes and to have extended lives *in vivo* and *in vitro*. PNAs also bind stronger to a complimentary DNA sequence due to the lack of charge repulsion between the PNA strand and the DNA strand. Other nucleic acid analogs may contain nucleotides containing polymer backbones, cyclic backbones, or acyclic backbones. For example, the nucleotides may have morpholino backbone structures (U.S. Pat. No. 5,034,506). The analogs may also contain groups such as reporter groups, a group for improving the pharmacokinetic or pharmacodynamic properties of nucleic acid sequence.

Other analogs include isolated nucleic acid sequences having sequences which differ from the nucleic acid sequence shown in Figure 1 (SEQ.ID.NO.:1) due to degeneracy in the genetic code are also within the

scope of the invention. Such nucleic acids encode functionally equivalent proteins but differ in sequence from the above mentioned sequences due to degeneracy in the genetic code.

The present invention further includes the preparation or isolation of other nucleic acid sequences which are the same, analogous, homologous or can hybridize to the nucleic acid sequences of the invention. For example, an isolated nucleic acid molecule of the invention which comprises DNA can be isolated by preparing a labelled nucleic acid probe based on all or part of the nucleic acid sequences as shown in Figure 1 (SEQ.ID.NO.:1), and using this labelled nucleic acid probe to screen an appropriate DNA library (e.g. a cDNA or genomic DNA library). For example, a genomic library can be used to isolate a DNA by screening the library with the labelled probe using standard techniques. Nucleic acids isolated by screening of a cDNA or genomic DNA library can be sequenced by standard techniques.

An isolated nucleic acid molecule of the invention which is DNA can also be isolated by selectively amplifying a nucleic acid using the polymerase chain reaction (PCR) methods and cDNA or genomic DNA. It is possible to design synthetic oligonucleotide primers from the nucleic acid sequence as shown in Figure 1 (SEQ.ID.NO.:1), for use in PCR. A nucleic acid can be amplified from cDNA or genomic DNA using these oligonucleotide primers and standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. It will be appreciated that cDNA may be prepared from mRNA, by isolating total cellular mRNA by a variety of techniques, for example, by using the guanidinium-thiocyanate extraction procedure of Chirgwin et al., Biochemistry, 18, 5294-5299 (1979). cDNA is then synthesized from the mRNA using reverse transcriptase (for example, Moloney MLV reverse transcriptase available from Gibco/BRL, Bethesda, MD, or AMV reverse transcriptase available from Seikagaku America, Inc., St. Petersburg, FL).

An isolated nucleic acid molecule of the invention which is RNA can be isolated by cloning a cDNA of the invention into an appropriate vector which allows for transcription of the cDNA to produce an RNA molecule. For example, a cDNA can be cloned downstream of a bacteriophage promoter, (e.g., a T7 promoter) in a vector, cDNA can be transcribed *in vitro* with T7 polymerase, and the resultant RNA can be isolated by standard techniques.

A nucleic acid molecule of the invention may also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071).

The sequence of a nucleic acid molecule of the invention may be inverted relative to its normal presentation for transcription to produce an antisense nucleic acid molecule. Preferably, an antisense sequence is constructed by inverting a region preceding the initiation codon or an unconserved region. In particular, the nucleic acid sequences contained in the nucleic acid molecules of the invention or a fragment thereof, preferably a nucleic acid sequence shown in Figure 1 (SEQ.ID.NO.:1) may be inverted relative to its normal presentation for transcription to produce antisense nucleic acid molecules.

The antisense nucleic acid molecules of the invention or a fragment thereof, may be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed with mRNA or the native gene e.g. phosphorothioate derivatives and acridine substituted nucleotides. The antisense sequences may be produced biologically using an expression vector introduced into cells in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense sequences are produced under the control of a high

efficiency regulatory region, the activity of which may be determined by the cell type into which the vector is introduced.

## II. USES OF THE MARKER

As previously mentioned, the isolation of the novel marker for E. coli serotypes O157:H7; O157:NM and O55:H7 allows the development of diagnostic assays that can be used to detect the serotypes in a sample. The sample can be any sample, including but not limited to, clinical, food, water and environmental samples. Clinical samples include bodily materials such as blood, urine, serum, tears, saliva, feces, tissues and the like.

Accordingly, the present invention provides a method of detecting the presence of E. coli serotypes O157:H7; O157:NM or O55:H7 in a sample comprising (a) isolating nucleic acid from the sample and (b) determining if the isolated nucleic acid contains (1) a nucleic sequence shown in Figure 1 (SEQ.ID.NO.:1) wherein T can also be U or a diagnostic fragment thereof; (2) a nucleic sequence that is complimentary to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (3) a nucleic sequence that can hybridize to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (4) a nucleic acid sequence which has substantial sequence homology to (1), (2) or (3); or (5) a nucleic acid sequence which is an analog of any of the nucleic acid sequences of (1) to (4), wherein the presence of a nucleic acid sequence defined in (1) to (5) indicates the presence of E. coli serotypes O157:H7; O157:NM or O55:H7 in the sample.

### (a) Primers

The present invention includes the preparation of nucleic acid primers based on the sequence of the marker shown in Figure 1 (SEQ.ID.NO.:1). Accordingly, the present invention provides an isolated nucleic acid primer comprising a portion of a (1) a nucleic sequence shown in Figure 1 (SEQ.ID.NO.:1) wherein T can also be U or a diagnostic fragment thereof; (2) a nucleic sequence that is complimentary to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment

thereof; (3) a nucleic sequence that can hybridize to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (4) a nucleic acid sequence which has substantial sequence homology to (1), (2) or (3); or (5) a nucleic acid sequence which is an analog of any of the nucleic acid sequences of (1) to (4). Preferably, the primer contains from about 5 to about 50 nucleotides, more preferably from about 15 to 30 nucleotides.

The length and bases of primers are selected so that they will hybridize to different strands of the sequence as shown in Figure 1 (SEQ.ID.NO.:1) and at relative positions along the sequence such that an extension product synthesized from one primer when it is separated from its template can serve as a template for extension of the other primer into a nucleic acid of defined length. Primers which may be used in the invention are oligonucleotides, i.e., molecules containing two or more deoxyribonucleotides of the nucleic acid molecule of the invention which occur naturally as in a purified restriction endonuclease digest or are produced synthetically using techniques known in the art such as for example phosphotriester and phosphodiester methods (See Good *et al.* Nucl. Acid Res 4:2157, 1977) or automated techniques (See for example, Conolly, B.A. Nucleic Acids Res. 15:15(7): 3131, 1987). The primers are capable of acting as a point of initiation of synthesis when placed under conditions which permit the synthesis of a primer extension product which is complementary to a DNA sequence of the invention, i.e., in the presence of nucleotide substrates, an agent for polymerization such as DNA polymerase and at suitable temperature and pH. Preferably, the primers are sequences that do not form secondary structures by base pairing with other copies of the primer or sequences that form a hair pin configuration.

It will be appreciated that the primers may contain non-complementary sequences provided that a sufficient amount of the primer contains a sequence which is complementary to a nucleic acid molecule of the invention or oligonucleotide fragment thereof, which is to be amplified. Restriction site linkers may also be incorporated into the

primers allowing for digestion of the amplified products with the appropriate restriction enzymes facilitating cloning and sequencing of the amplified product.

In one embodiment, the primer comprises nucleotides 597-618 of the sequence shown in Figure 1 (SEQ.ID.NO.:1). In another embodiment, the primer comprises the reverse complement of nucleotides 1136-1155 of the sequence shown in Figure 1 (SEQ.ID.NO.:1). In a specific embodiment, the primer is O157-F and has the sequence 5'-CGGTTTAATGGCTTGTTGTGCT-3' (SEQ.ID.NO.:3). This primer corresponds to the sequence at nucleotides 597-618 of SEQ.ID.NO.:1 except that the G residue at position 617 has been replaced with C in the primer. In another specific embodiment, the primer is O157-R and has the sequence 5'-ATGCCATTAAACCGGTGGC-3' (SEQ.ID.NO.:4). This primer corresponds to the reverse complement of the nucleotides found at positions 1136-1155 of SEQ.ID.NO.:1. These are shown in Table 3.

Using the primers illustrated in Table 3 in a PCR assay, the inventors have demonstrated that these primers are specific for the E. coli serotypes O157:H7; O157:NM and O55:H7 but are not specific for 119 other E. coli strains belonging to 60 serotypes and 59 isolates belonging to 44 non-E. coli species (see Example 1).

Accordingly, the present invention provides a method of detecting the presence of E. coli serotypes O157:H7; O157:NM and O55:H7 in a sample comprising (a) isolating nucleic acid from the sample; (b) amplifying the isolated nucleic acid with a primer comprising a sequence that is complimentary to a portion of (1) a nucleic sequence shown in Figure 1 (SEQ.ID.NO.:1) wherein T can also be U or a diagnostic fragment thereof; (2) a nucleic sequence that is complimentary to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (3) a nucleic sequence that can hybridize to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (4) a nucleic acid sequence which has substantial sequence homology to (1), (2) or (3); or (5) a nucleic acid sequence which is an analog of any of the nucleic acid

sequences of (1) to (4); and (c) assaying for amplified sequences, wherein the presence of an amplified sequence indicates that the sample contains one or more of E. coli serotypes O157:H7; O157:NM and O55:H7. Preferably, a primer pair is used, one with a sequence shown in SEQ.ID.NO.:3 and the other with a sequence shown in SEQ.ID.NO.:4.

The nucleic acid sequences may be amplified in step (b) using any method that results in the amplification of nucleic acid molecules such as the Polymerase Chain Reaction, Ligase Chain Reaction or NASBA may be used to amplify a nucleic acid molecule of the invention (Barney in "PCR Methods and Applications", August 1991, Vol.1(1), page 5, and European Published Application No. 0320308, published June 14, 1989, and U.S. Serial No. 5,130,238 to Malek).

Preferably, the nucleic acid sequences are amplified in step (b) using a Polymerase Chain Reaction (PCR). The conditions which may be employed in the methods of the invention using PCR are those which permit hybridization and amplification reactions to proceed in the presence of DNA in a sample and appropriate complementary hybridization primers. Conditions suitable for the polymerase chain reaction are generally known in the art. For example, see M.A. Innis and D.H. Gelfand, PCR Protocols, A guide to Methods and Applications M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White eds, pp3-12, Academic Press 1989, which is incorporated herein by reference. Preferably, the PCR utilizes polymerase obtained from the thermophilic bacterium *Thermus aquatics* (Taq polymerase, GeneAmp Kit, Perkin Elmer Cetus) or other thermostable polymerase may be used to amplify DNA template strands.

The amplified products can be isolated and distinguished based on their respective sizes using techniques known in the art. For example, after amplification, a DNA sample can be separated on an agarose gel and visualized, after staining with ethidium bromide, under ultra violet (uv) light. DNA may be amplified to a desired level and a further extension reaction may be performed to incorporate nucleotide derivatives having detectable markers such as radioactive labelled or biotin labelled

nucleoside triphosphates. The primers may also be labelled with detectable markers as discussed below. The detectable markers may be analyzed by restriction and electrophoretic separation or other techniques known in the art.

The primers may be labelled with detectable markers which allow for detection of the amplified products. Suitable detectable markers are radioactive markers such as P-32, S-35, I-125, and H-3; luminescent markers such as chemiluminescent markers, preferably luminol; and fluorescent markers, preferably dansyl chloride, fluorcein-5-isothiocyanate, and 4-fluor-7-nitrobenz-2-axa-1,3 diazole, Cy3, Cy5, Texas Red, NED; enzyme markers such as horseradish peroxidase, alkaline phosphatase, β-galactosidase, acetylcholinesterase, or biotin.

**(b) Probes**

The present invention also includes the preparation of nucleic acid probes based on the sequence of the marker shown in Figure 1 (SEQ.ID.NO.:1). Accordingly, the present invention provides an isolated nucleotide probe comprising a portion of (1) a nucleic sequence shown in Figure 1 (SEQ.ID.NO.:1) wherein T can also be U or a diagnostic fragment thereof; (2) a nucleic sequence that is complimentary to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (3) a nucleic sequence that can hybridize to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (4) a nucleic acid sequence which has substantial sequence homology to (1), (2) or (3); or (5) a nucleic acid sequence which is an analog of any of the nucleic acid sequences of (1) to (4).

Preferably, the probe comprises a portion of the sequence shown in Figure 1 (SEQ.ID.NO.:1). More preferably, the probe comprises from about 20 to about 500 nucleotides, even more preferably 50 to 200 nucleotides from the sequence shown in Figure 1 (SEQ.ID.NO.:1).

In one embodiment, the probe comprises nucleotides 597-677 of the sequence shown in Figure 1 (SEQ.ID.NO.:1). Preferably, the probe is O157-A and has the sequence CGGTTTAATGGCTTGTTGTGGTAA

CACCGAAGCCAGCTCAATAAATTGCTGCGATGAGTTAC AGCTATCGAGTAAACCACC (SEQ.ID.NO.:5).

In another embodiment, the probe comprises nucleotides 1059-1155 of Figure 1 (SEQ.ID.NO.:1). Preferably, the probe is O157-B and has the sequence TATCCCGAATCGCCTGGCGTTTTTGCACATC CTCTGCGACGCTATTTTTGTGGAACGCAAAGCCATCAAGGAAAATA GCCA CCGGTTTAAT GGCAT (SEQ.ID.NO.:6).

The probes can be used to detect the presence or absence of E. coli serotypes O157:H7; O157:NM and O55:H7 in a sample.

Accordingly, the present invention also relates to a method of detecting the presence of a nucleic acid molecule associated with E. coli serotypes O157:H7; O157:NM and O55:H7 in a sample comprising (a) contacting the sample under hybridization conditions with one or more nucleotide probes which hybridize to the nucleic acid molecules and (b) determining if there is hybridization between the nucleic acid molecules in the sample and the nucleotide probes wherein the presence of hybridization indicates that the sample contains one of E. coli serotypes O157:H7; O157:NM and O55:H7.

Hybridization conditions which may be used in the methods of the invention are known in the art and are described for example above in the definition of "nucleic acid sequences that hybridize". The hybridization product may be assayed using techniques known in the art. The nucleotide probe may be labelled with a detectable marker and the hybridization product may be assayed by detecting the detectable marker or a detectable change produced by the detectable marker.

The detectable marker used to label the probe can be any marker such as a radioactive label which provides for an adequate signal and has sufficient half life such as $^{32}P$, $^{3}H$, $^{14}C$ or the like. Other detectable markers which may be used include antigens that are recognized by a specific labelled antibody, fluorescent compounds, enzymes, antibodies specific for a labelled antigen, and chemiluminescent compounds. An appropriate label may be selected having regard to the rate of hybridization

and binding of the probe to the nucleotide to be detected and the amount of nucleotide available for hybridization.

**(c) Microchips**

The present invention also includes microchips comprising (a) an isolated nucleic acid molecule comprising (1) a nucleic sequence shown in Figure 1 (SEQ.ID.NO.:1) wherein T can also be U or a diagnostic fragment thereof; (2) a nucleic sequence that is complimentary to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (3) a nucleic sequence that can hybridize to the sequence shown in Figure 1 (SEQ.ID.NO.:1) or a diagnostic fragment thereof; (4) a nucleic acid sequence which has substantial sequence homology to (1), (2) or (3); or (5) a nucleic acid sequence which is an analog of any of the nucleic acid sequences of (1) to (4) attached to (b) a microchip.

In one embodiment, the microchip comprises a probe having the sequence CGGTTTAATGGCTTGTTGTGGTAA CACCGAAGCCAGCTCAATAAATTGCTGCGATGAGTTAC AGCTATCGAGTAAACCACC (SEQ.ID.NO.:5) attached to a microchip.

In another embodiment, the microchip comprises a probe having the sequence TATCCCGAATCGCCTGGCGTTTTTGCACATC CTCTGCGACGCTATTTTTGTGGAACGCAAAGCCATCAAGGAAAATA GCCA CCGGTTTAAT GGCAT (SEQ.ID.NO.:6) attached to a microchip.

As described in Example 2, microchips containing the probes shown in SEQ.ID.NO.:5 and SEQ.ID.NO.:6 were useful in detecting E. Coli serotype O157:H7.

**(d) Kits**

Reagents suitable for conducting the above described diagnostic methods of the invention may be packaged into convenient kits providing the necessary materials, packaged into suitable containers. Such kits may include all the reagents required to detect an E. coli serotype in a sample by means of the methods described herein, such as appropriate probes and/or primers for performing PCR.

In one embodiment of the invention the kit contains a nucleotide probe which hybridizes with a nucleic acid molecule of the invention, reagents required for hybridization of the nucleotide probe with the nucleic acid molecule, and directions for its use. In another embodiment of the invention, the kit includes primers which are capable of amplifying a nucleic acid molecule of the invention or a predetermined oligonucleotide fragment thereof, all the reagents required to produce the amplified nucleic acid molecule or predetermined fragment thereof in the polymerase chain reaction, and means for assaying the amplified sequences.

As mentioned above, the methods and kits of the present invention have many practical applications. For example, the methods and kits of the present invention may be used to detect E. Coli serotypes 0157:H7, 0157:NM or 055:H7 in any sample suspected of containing E. coli. Samples which may be tested include bodily materials such as blood, urine, serum, tears, saliva, feces, tissues and the like. Further, water and food samples and other environmental samples and industrial wastes may be tested.

Before testing a sample in accordance with the methods described herein, the sample may be concentrated using techniques known in the art, such as centrifugation and filtration. For hybridization and/or PCR-based methods described herein, nucleic acids may be extracted from cell extracts of the test sample using techniques known in the art such as precipitation, solvent extraction and column purification.

The following non-limiting examples are illustrative of the present invention:

## EXAMPLES

### Example 1

*AFLP marker identification*

The marker was initially identified as a 360 bp fluorescent fragment (Figure 2 or SEQ.ID.NO.:2) using the FAFLP method (Vos et al., 1995. Nucleic Acids Research, Vol. 23, pp. 4407-4414). The FAFLP analysis

was performed by the selective amplification of restriction fragments from a digest of total genomic DNA with restriction enzymes *EcoR* I and *Mse* I. The DNA sequences of the selective primer pair that produced the polymorphic marker are as shown in Table 1 and in SEQ.ID.NOS:7 and 8. Among 163 *E. coli* strains tested that belonged to 40 serotypes (Table 2), only serotypes O157:H7, O157:NM, O145:NM and O55:H7 contained the 360 bp fragment.

***DNA sequencing***

The target fragment was directly isolated from the AFLP gel using ABI 377 automated DNA Sequencer and re-amplified by polymerase chain reaction (PCR), and then sequenced. The DNA sequence of the 360 bp fragment is shown in Figure 2 (SEQ.ID.NO.:2). A 1223 bp downstream region continued from the AFLP fragment was further sequenced using a single primer walking method with strains of *E. coli* serotypes O157:H7, O157:NM, O55:H7 and O145:NM. Four mutations (C-T, A-G, T-A and G-A) in the region were identified in the strain of serotype O145:NM. The complete sequence of the 1583 bp fragment is shown in Figure 1 (SEQ.ID.NO.:1). The result of BLAST search showed that the DNA sequences of the 1583 bp fragment had no significant homology to 400 sequences of the *E. coli* genome and 400,635 known sequences in the GenBank, EMBL, DDBJ, PDB DNA databases. The DNA sequences of the 1583 bp fragment were also analyzed by a software program, GeneWorks 2.5; a potential open reading frame (ORF) was identified within the sequences of 1043 - 90 nt.

***Marker application***

The DNA fragment has been useful as a marker in developing a PCR assay for specific detection of *E. coli* serotypes O157:H7, O157:NM and O55:H7. The DNA sequences of the primers used in the PCR assay are provided in Table 3. Primer O157-F contains one mismatch to the sequences of O157:H7/O157:NM for eliminating possible cross reaction from the sequences of O145:NM. The primer pair results in a 560 bp PCR product under the following conditions. An amplification

reaction mixture (20 µL) contained 10mM Tris-HCl and 50 mM Kcl (pH 8.3), 1.2 mM $MgCl_2$, 200 µM each of dATP, dCTP, dGTP and dTTP, 0.4 pmol/µL each of the primers, 0.5 unit of *Taq* DNA polymerase (Perkin-Elmer) and 5 ul of DNA template. The thermal cycling conditions (GeneAmp 9600 PCR System, Perkin-Elmer) were as follows: initial denaturation at 94°C for 3 min; 35 cycles consisting of denaturation at 94°C for 30 sec, annealing at 60°C for 30 sec and extension at 72°C for 45 sec; incubation at 72°C for 5 min. PCR products were visualized on a 1.6% agarose gel after staining with ethidium bromide.

The specificity of the primers and the PCR assay has been evaluated with a total of 119 *E. coli* strains belonging to 60 serotypes (Table 4) and 59 isolates belonging to 44 non-*E. coli* species (Table 5). The 560 bp specific PCR product appeared only with strains of serotypes O157:H7, O157:NM and O55:H7 (Figure 3) but not the strains of the other serotypes of E. coli or other bacterial strains tested. The PCR assay can be used for the detection of *E.coli* O157:H7 in food, clinical, water and environmental samples.

**Example 2**

**Preparation of a Microchip**

**MATERIALS AND METHODS**

***DNA microchip system***

The arrayer (with operation software) was a Brown Lab/Stanford System that was assembled at Virtek Vision (Waterloo, Ontario) and can produce 1000 to 2500 spots/$cm^2$. The DNA scanner, ChipReader™ (with data collection software) was from Virtek Vision. The software ImaGene was the product of BioDiscovery and contributed by Virtek.

***DNA probe preparation***

In this Example, "probe" refers to the DNA covalently attached to the chip surface and "target" in the fluorescent molecules that hybridize to this probe.

The probes for *E. coli* serotype O157:H7 were designed according to the sequences of the O157:H7 marker and are illustrated in Table 6. Short (about 20 bp) probes were prepared by direct synthesis and 5'-end labeling with amine. Other probes were prepared by PCR using the DNA samples from the control *E. coli* strains and specific primers with one of them labeled with 5'-amine.

PCR reaction mixtures (50μL) contain 1x amplification Buffer II (Perkin-Elmer), 200 μM dNTPs, 2 mM $mgCl_2$, 0.5 μM forward and reverse primers, 1 unit of AmpliTag DNA polymerase (Perkin-Elmer) and 10μL of template DNA. Amplification for probe preparation consisted of incubation at 95°C for 1 min. 32 cycles of 30s at 94°C, 30s at 52°C and 1 min at 72°C, and extension at 72°C for 7 min. PCR products were visualized using ethidium bromide-stained agarose gels, purified using QIAquick PCR purification Kit (Qiagen) and dissolved in 3x SSC buffer for spotting.

***Probe spotting and immobilization on the chips***

The purified DNA probes with amine-modified tags were spotted on a glass microscope slide that is coated with aldehyde groups (Telechem International, Inc.). The DNA probes were immobilized onto the glass surface by covalent binding via the Schiff's base attachment. Spotting was carried out with the Stanford arrayer system. This robot system is designed to automatically collect samples (DNA probes) from either 96- or 384-well microtitre plates, and to spot the samples onto up to 16 slides simultaneously. The probes were spotted with a single pin at a density of 2500 spots/$cm^2$ (200 μm spacing). After spotting, the slides were rehydrated in a humidity chamber for 2-4 h, and rinsed once with 0.2% SDS and twice with water. The slides were then incubated for 5 min in a sodium borohydride solution (1.0%). The probes (if they were PCR products) were denatured by heating the slides at 95°C for 2 min, rinsing once with 0.2% SDS and twice with water. The slides were dried and stored at room temperature until use.

*Fluorescent target DNA preparation*

Fluorescent target DNAs were prepared by PCR from the control strains of bacteria with specific primers. The PCR reaction conditions were same as those used in the probe preparation except the following modifications: the primers were not labeled; 50% of the normal dCTP was substituted with 500 µM of fluorescent Cy3-dCTP (Amersham Pharmacia Biotech) in the PCR reaction mixtures. The labeled target DNAs were purified with MicroSpin columns (Amersham Pharmacia Biotech) and stored at -20°C until use.

*DNA hybridization*

The immobilized probes were pre-hybridized with 1 µg of salmon sperm DNA in 20 µL of the DIG Easy Hyb solution (Boehringer Mannheim) at 65°C (37°C for small fragments) for 1 h. The target DNAs (10 µL) were denatured at 95°C for 2 min, and mixed with 10 µL of the DIG Easy Hyb solution and applied onto the processed slide with a coverslip. The slides were placed in a humid chamber and incubated overnight at 37°C or 62°C. Following hybridization, the slides were washed with the washing buffers with different stringency for four times and dried at room temperature.

*Data collection and analysis*

Hybridization patterns were scanned using the ChipReader™ (Virtek). Data were collected using the ChipReader™ collection software (Virtek) at 100% laser intensity, 700-900 detector sensitivity, 2 detector gain on the Cy-3 channel. Fluorescent images were analyzed using the ImaGene™ software (BioDiscovery).

*Results*

The marker of the invention has been demonstrated to be useful in developing a DNA microarray/chip-based assay for specific detection of the *E. coli* serotypes O157:H7, O157:NM and O55:H7 in a sample. Table 6 shows the sequences of the two DNA probes that were designed based on the marker. The DNA probes suitable for spotting were

generated using PCR and tagged with amine at their 5′ end. The probes were used together with other 22 probes and immobilized onto the microscopic slide surface by covalent binding via Schiff's base attachment. The microarray was used in hybridization with Cy3 labeled target DNA prepared from 10 different reference strains of *E. coli*. The inventors have shown that the two specific probes on the chip specifically hybridized to the strains of *E. coli* serotype O157:H7.

While the present invention has been described with reference to what are presently considered to be the preferred examples, it is to be understood that the invention is not limited to the disclosed examples. To the contrary, the invention is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

All publications, patents and patent applications are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

**Table 1**

**DNA sequences of the selective primers that produced the AFLP marker**

| Primer | Oligo-nucleotide sequence | |
|---|---|---|
| EcoR I-C | 5'-GACTGCGTACCAATTCC-3' | (SEQ.ID.NO.:7) |
| Mse I-G | 5'-GATGAGTCCTGAGTAAG-3' | (SEQ.ID.NO.:8) |

**Table 2**

**E. coli strains screened for the presence of the O157:H7 marker in AFLP analysis**

| Serotype | No. of strains | Serotype | No. of strains | Serotype | No. of strains |
|---|---|---|---|---|---|
| O?:H7 | 1 | O91:H21 | 5 | O139:K82 | 1 |
| O?:H8 | 1 | O91:NM | 2 | O145:NM | 2 |
| O5:NM | 4 | O98:NM | 1 | O153:H25 | 3 |
| O7:H4 | 4 | O103:H2 | 24 | O153:H31 | 1 |
| O8:H19 | 1 | O111:H8 | 1 | O156:H7 | 1 |
| O22:H8 | 5 | O111:NM | 5 | O156:NM | 1 |
| O26:H11 | 6 | O113:H4 | 1 | O157:H7 | 56 |
| O39:H49 | 1 | O113:H21 | 2 | O157:H19 | 1 |
| O46:H38 | 1 | O115:H8 | 1 | O157:H25 | 1 |
| O55:H7 | 3 | O118:H16 | 1 | O157:NM | 2 |
| O76:H25 | 1 | O121:H7 | 1 | O163:H19 | 1 |
| O80:NM | 3 | O127:H6 | 1 | O163:NM | 1 |
| O88:H25 | 1 | O128:B12 | 1 | | |
| O91:H14 | 3 | O132:NM | 11 | | |

**Table 3**

**DNA sequences of the primers specific for**

**E. coli O157:H7, O157:NM and O55:H7**

| Primer | Oligo-nucleotide sequence | Location within the marker sequence |
|---|---|---|
| O157-F (22-mer) | 5'-CGGTTTAATGGCTTGTTGTGCT-3' | 597 - 618 (SEQ.ID.NO.:3) |
| O157-R (19-mer) | 5'-ATGCCATTAAACCGGTGGC-3' | 1136 - 1155 (SEQ.ID.NO.:4) |

**Table 4**

**E. coli strains tested for the presence of the O157:H7 marker in the PCR assay**

| Serotype | No. of strains | Serotype | No. of strains | Serotype | No. of strains |
|---|---|---|---|---|---|
| O?:H2 | 1 | O55:H7 | 6 | O126:H8 | 1 |
| O?:H7 | 1 | O76:H25 | 1 | O127:H6 | 1 |
| O?:H8 | 1 | O80:NM | 2 | O128:B12 | 1 |
| O?:H19 | 1 | O84:H2 | 1 | O132:NM | 2 |
| O?:H21 | 1 | O88:H25 | 1 | O136:H16 | 1 |
| O2:H29 | 1 | O91:H14 | 3 | O139:K82 | 1 |
| O5:NM | 2 | O91:H21 | 3 | O142:H38 | 1 |
| O7:H4 | 2 | O91:NM | 2 | O145:NM | 5 |
| O8:H? | 1 | O98:NM | 1 | O153:H25 | 3 |
| O8:H9 | 1 | O103:H2 | 5 | O153:H31 | 1 |
| O8:H19 | 3 | O111:H8 | 2 | O153:NM | 1 |
| O15:H27 | 1 | O111:NM | 5 | O156:H7 | 1 |
| O15:NM | 1 | O113:H4 | 1 | O156:NM | 1 |
| O22:H8 | 2 | O113:H21 | 2 | O157:H7 | 15 |
| O26:H11 | 5 | O115:H8 | 1 | O157:H19 | 1 |
| O26:NM | 1 | O116:H21 | 1 | O157:H25 | 2 |
| O38:H21 | 1 | O118:H16 | 1 | O157:NM | 6 |
| O39:H49 | 1 | O121:H6 | 1 | O163:H19 | 1 |
| O45:H2 | 1 | O121:H7 | 2 | O163:NM | 2 |
| O46:H38 | 1 | O121:H19 | 1 | O165:NM | 1 |

**Table 5**

**Non-E. coli species tested for the presence of the O157:H7 marker in the PCR assay**

| Species | No. of isolates |
|---|---|
| Acinetobacter calcoaceticus | 1 |
| Aeromonus hydrophila | 2 |
| Bacillus cereus | 2 |
| Bacillus circuluris | 1 |
| Bacillus parabrevis | 1 |
| Bacillus subtilis | 1 |
| Citrobacter diversus | 1 |
| citrobacter freudii | 1 |
| Campylobacter | 2 |
| Campylobacter coli | 2 |
| Campylobacter jejuni | 1 |
| Canadida albicus | 1 |
| Enterobacter cloacae | 1 |
| Enterobacter amnigenus | 1 |
| Enterococcus fecalis | 2 |
| Edwardsiella tarda | 1 |
| Kluyvera ascovbata | 1 |
| Listeria monocytogenes | 1 |
| Listeria ivarovii | 1 |
| Lactobacillus brevis | 1 |
| Lactobacillus planterum | 1 |
| Moraxella osloensis | 1 |
| Proteus vulgaris | 1 |
| Proteus mirabilis | 1 |
| Pseudomonus aeruginosa | 1 |
| Salomonella tryphimurium | 4 |
| Salmonella heidelberg | 1 |
| Salmonella thompson | 1 |
| Salmonella newport | 1 |
| Salmolella hadar | 2 |
| Salmonella infantis | 2 |
| Salmonella schwarzengrund | 2 |
| Salmonella choleraesuis var.kunzendorf | 2 |
| Shigella flexaeri | 1 |
| Shigella sonnei | 1 |
| Straphylococcus aureus | 4 |
| Straphylococcus epidermidis | 1 |
| Straphylococcus haemolyticus | 1 |
| Straphylococcus lugdenersis | 1 |
| Straphylococcus intermedium | 1 |
| Streptococcus pyogenes | 1 |
| Streptococcus bovis | 1 |
| Xanthomonas malitophilia | 1 |
| Yersinia enterocolitica | 1 |

**Table 6**
**DNA sequences of the probes specific for E. coli O157:H7, O157:NM and O55:H7**

| Probe | Nucleotide sequence | Location within the marker sequence |
|---|---|---|
| O157-A | CGGTTTAATGGCTTGTTGTG GTAACACCGAAGCCAGCTCA ATAAATTGCTGCGATGAGTT ACAGCTATCGAGTAAACCACC | 597 - 677 |
| O157-B | TATCCCGAATCGCCTGGCGT TTTTGCACATC CTCTGCGAC GCTATTTTTGTGGAACGCAA AGCCATCAAGGAAAATAGCC ACCGGTTTAATGGCAT | 1059 - 1155 |